Estimation adaptative pour des processus de comptage dépendant de covariables

Fabienne Comte[(1)] & Stéphane Gaïffas[(2)] & Agathe Guilloux[(2)]

[(1)] *MAP5, Université Paris Descartes, France.*

[(2)] *LSTA Université Pierre et Marie Curie, France.*

Abstract
We propose in this work an original estimator of the conditional intensity of a marker-dependent counting process, with the example of conditional hazard rate in mind. We use model selection methods and provide a non asymptotic risk bound for our estimator with respect to the integrated mean square risk on a compact set. This implies an asymptotic upper bound for the rate of convergence of the estimator, which is reached automatically on a functional class with given (unknown) anisotropic regularity.
Résumé
Nous proposons un nouvel estimateur de l'intensité conditionnelle de processus de comptage dépendant de covariables. Nous utilisons les méthodes de sélection de modèle et obtenons une borne non-asymptotique pour le MISE pour notre estimateur. Cette borne est atteinte automatiquement sur des classes fonctionnels de régularité anisoptrope inconnue. Nous montrons par ailleurs que cette borne est optimale au sens minimax.

1 Introduction

Soit N un processus de comptage dépendant de covariables, dont le compensateur Λ vérifie le modèle d'intensité multiplicative d'Aalen :

$$\Lambda(t) = \int_0^t \alpha(X, z) Y(z) dz, \text{ pour tout } t \geq 0 \tag{1}$$

où X est un vecteur de covariables dans $\mathbb{R}^d$, le processus Y est positif et prévisible et α est une fonction déterministe inconnue, appelée intensité. Nous souhaitons estimer l'intensité α (et des fonctionnelles de cette intensité) à partir de l'observation d'un n-échantillon $(X_i, N^i(z), Y^i(z), z \leq \tau)$ pour $i = 1, \ldots, n$ où $\tau < +\infty$.

Nous citons trois papiers dans lesquels différentes stratégies d'estimation de l'intensité de processus de comptage ont été proposées. Aucun d'entre eux ne considère la présence de covariables. Ramlau-Hausen (1983) a proposé un estimateur par noyau, Grégoire (1993) a étudié la validation croisée. Plus récemment, Reynaud-Bouret (2006) a considéré l'estimation adaptative par projection.

Exemple 1 *Modèle de régression pour des données censurées*
Soit T une variable aléatoire (v.a.) positive et X un vecteur de covariables dans $\mathbb{R}^d$, dont

les fonctions de répartition (f.d.r.) respectives sont notées F_T and F_X. Considérons que T peut être censurée par une v.a. C de f.d.r. G, de telle façon que les v.a. sont $Z = T \wedge C$, $\delta = \mathbb{1}(T \leq C)$ et X sont observables. Nous supposons que

$$(\mathcal{C}) : \ T \text{ et } C \text{ sont indépendantes conditionnellement à } X.$$

Dans ce cas, les processus à considérer sont donnés (voir e.g. Andersen et al. (1993)) pour $i = 1, \ldots, n$ et $z \geq 0$ par:

$$N^i(z) = \mathbb{1}(Z_i \leq z, \delta_i = 1) \text{ et } Y^i : Y^i(z) = \mathbb{1}(Z_i \geq z).$$

L'intensité inconnue α est ici égale au risque instantané conditionnel de T sachant $X = x$ défini, pour tout $z > 0$, par :

$$\alpha(x, z) = \alpha_{T|X}(x, z) = \frac{f_{T|X}(x, z)}{1 - F_{T|X}(x, z)},$$

où $f_{T|X}$ et $F_{T|X}$ sont resp. la densité et la f.d.r. conditionnelles de T sachant X.

L'estimation non-paramétrique du risque instantané conditionnel a été initiée par Beran (1981). Stute (1986), Dabrowska (1987), McKeague and Utikal (1995) et Li and Doss (1995) ont étendu ses résultats.

Nous permettons ici que la loi de la censure dépende des covariables X, comme dans Dabrowska (1989) et Heuchenne and Van Keilegom (2006).

2 Description de la procédure d'estimation

2.1 Definition du contraste

Soit $A = A_1 \times A_2$ un ensemble compact sur $\mathbb{R} \times \mathbb{R}_+$ sur lequel α sera estimée. On pose $A = [0, 1] \times [0, 1]$, sans perte de généralité, et, en particulier, $\tau = 1$. Soit h une fonction de $(L^2 \cap L^\infty)(A)$. Définissons le contraste d'estimation par

$$\gamma_n(h) = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 h^2(X_i, z) Y^i(z) dz - \frac{2}{n} \sum_{i=1}^{n} \int_0^1 h(X_i, z) dN^i(z). \tag{2}$$

Ce contraste de type moindres-carrés est adapté au problème considéré. Comme chaque N^i admet une décomposition de Doob-Meyer ($N^i = \Lambda^i + M^i$), on peut également écrire :

$$\gamma_n(h) = \frac{1}{n} \sum_{i=1}^{n} \int_0^1 h^2(X_i, z) Y^i(z) dz - \frac{2}{n} \sum_{i=1}^{n} \int_0^1 h(X_i, z) d\Lambda^i(z) - \frac{2}{n} \sum_{i=1}^{n} \int_0^1 h(X_i, z) dM^i(z),$$

ainsi :

$$\mathbb{E}\big(\gamma_n(h)\big) = \mathbb{E}\big(\int_0^1 h^2(X, z) Y(z) dz\big) - \mathbb{E}\big(2 \int_0^1 h(X, z) d\Lambda(z)\big).$$

Soit F_X la fonction de répartition des covariables X et $\|\cdot\|_\mu$ la norme définie par :

$$\|h\|_\mu^2 \;:=\; \mathbb{E}\Big(\int_0^1 h^2(X,z)Y(z)dz\Big) = \iint_A h^2(x,z)d\mu(x,z),$$

où $d\mu(x,z) := \mathbb{E}(Y(z)|X=x)F_X(dx)dz$. Le modèle d'intensité multiplicative d'Aalen, cf. Equation (1), nous assure que:

$$\mathbb{E}\big(\gamma_n(h)\big) = \|h\|_\mu^2 - 2\iint h(x,z)\alpha(x,z)\mathbb{E}(Y(z)|X=x)F_X(dx)dz = \|h-\alpha\|_\mu^2 - \|\alpha\|_\mu^2.$$

La dernière équation montre pourquoi minimiser $\gamma_n(\cdot)$ sur un ensemble de fonctions approprié est une stratégie d'estimation de α.

2.2 Hypothèses et notations

Nous introduisons ici les hypothèses et notations dont nous avons besoin par la suite. Définissons les normes

$$\|h\|^2 = \iint h^2(x,z)dxdz, \|h\|_A^2 = \iint_A h^2(x,z)dxdz \text{ et } \|h\|_{\infty,A} = \sup_{(x,z)\in A}|h(x,z)|,$$

et nous supposons que les hypothèses suivantes sont vérifiées.

- $(\mathcal{A}1)$ Les covariables X_i admettent une densité f_X tel que $\sup_{A_1}|f_X| < +\infty$.

Cet hypothèse $(\mathcal{A}1)$ implique que μ admet une densité, notée f, par rapport à la mesure de Lebesgue.

- $(\mathcal{A}2)$ Il existe $f_0 > 0$, tel que $\forall (x,z) \in A_1 \times A_2,\ f(x,z) \geq f_0$.
- $(\mathcal{A}3)$ $\forall (x,z) \in A_1 \times A_2,\ \alpha(x,z) \leq \|\alpha\|_{\infty,A} < +\infty$.
- $(\mathcal{A}4)$ $\forall i, \forall t, Y^i(t) \leq C_Y$ où C_Y est une constante connue et fixée.

2.3 Définition de l'estimateur

Introduisons une collection $\{S_m, m \in \mathcal{M}_n\}$ d'espaces de projection : S_m est un modèle et $\mathcal{M}_n$ est un ensemble de multi-index. Pour tout $m = (m_1, m_2)$, S_m est un espace de fonctions à support dans $A = A_1 \times A_2$ défini par :

$$S_m = F_{m_1} \otimes H_{m_2} = \{h, \quad h(x,z) = \sum_{j\in J_m}\sum_{k\in K_m} a_{j,k}^m \varphi_j^m(x)\psi_k^m(z),\ a_{j,k}^m \in \mathbb{R}\},$$

où F_{m_1} et H_{m_2} sont des sous-espaces de $(L^2 \cap L^\infty)(\mathbb{R})$ engendrés par deux bases orthonormales $(\varphi_j^m)_{j\in J_m}$ avec $|J_m| = D_{m_1}$ et $(\psi_k^m)_{k\in K_m}$ avec $|K_m| = D_{m_2}$. Pour tout j et tout k, les supports de φ_j^m et ψ_k^m sont resp. inclus dans A_1 et A_2.

La première étape devrait être de définir $\hat{\alpha}_m = \text{argmin}_{h\in S_m} \gamma_n(h)$. Soit $h(x,y) = \sum_{j\in J_m}\sum_{k\in K_m} a_{j,k}\varphi_j^m(x)\psi_k^m(y)$ une fonction de S_m. Pour calculer $\hat{\alpha}_m$, il faut résoudre :

$$\forall j_0 \forall k_0, \quad \frac{\partial \gamma_n(h)}{\partial a_{j_0,k_0}} = 0 \Leftrightarrow G_m A_m = \Upsilon_m,$$

où A_m est la matrice $(a_{j,k})_{j\in J_m, k\in K_m}$,

$$G_m := \Big(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\varphi_l^m(X_i)\int \psi_k^m(z)\psi_p^m(z)Y^i(z)dz\Big)_{(j,k),(l,p)\in J_m\times K_m}$$

et

$$\Upsilon_m := \Big(\frac{1}{n}\sum_{i=1}^n \varphi_j^m(X_i)\int \psi_k^m(z)dN^i(z)\Big)_{j\in J_m, k\in K_m}.$$

Malheureusement G_m peut ne pas être inversible. Nous définissons donc $\hat{\alpha}_m$ de la manière suivante :

$$\hat{\alpha}_m := \begin{cases} \text{argmin}_{h\in S_m} \gamma_n(h) & \text{on } \hat{\Gamma}_m \\ 0 & \text{sur } \hat{\Gamma}_m^{\complement} \end{cases}, \tag{3}$$

où

$$\hat{\Gamma}_m := \Big\{ \min \text{Sp}(G_m) \geq \max(\hat{f}_0/3, n^{-1/2}) \Big\}$$

où $\text{Sp}(G_m)$ est le spectre de G_m. L'estimateur $\hat{f}_0$ de f_0 (cf. $(\mathcal{A}2)$) est supposé vérifier :

- $(\mathcal{A}5)$ pour tout k, $k \geq 1$, $\mathbb{P}(|\hat{f}_0 - f_0| > f_0/2) \leq C_k/n^k$.

Après avoir défini l'estimateur $\hat{\alpha}_m$, nous sélectionnons l'espace approprié via le critère :

$$\hat{m} = \underset{m\in\mathcal{M}_n}{\text{argmin}} \Big(\gamma_n(\hat{\alpha}_m) + \text{pen}(m)\Big). \tag{4}$$

Nous choisissons $\hat{\alpha}_{\hat{m}}$ comme estimateur de α sur A.

2.4 Hypothèses sur les modèles et exemples

Les hypothèses sur les modèles sont les suivantes :

- $(\mathcal{M}1)$ Pour $i = 1, 2$, $\mathcal{D}_n^{(i)} := \max_{m\in\mathcal{M}_n} D_{m_i} \leq n^{1/4}/\sqrt{\log n}$.
- $(\mathcal{M}2)$ Il existe des réels positifs ϕ_1, ϕ_2 tels que, pour tout u dans F_{m_1} et tout v dans H_{m_2}, nous avons

$$\sup_{x\in A_1} |u(x)|^2 \leq \phi_1 D_{m_1}\int_{A_1} u^2 \text{ et } \sup_{x\in A_2}|v(x)|^2 \leq \phi_2 D_{m_2}\int_{A_2} v^2.$$

En posant $\phi_0 = \sqrt{\phi_1\phi_2}$, on obtient

$$\forall h \in S_m \qquad \|h\|_{\infty,A} \leq \phi_0\sqrt{D_{m_1}D_{m_2}}\|h\|_A. \tag{5}$$

- $(\mathcal{M}3)$ Condition d'emboîtement. $D_{m_1} \leq D_{m'_1} \Rightarrow F_{m_1} \subset F_{m'_1}$ et $D_{m_2} \leq D_{m'_2} \Rightarrow H_{m_2} \subset H_{m'_2}$. Il existe un espace englobant global $\mathcal{S}_n$ dans la collection, tel que $\forall m \in \mathcal{M}_n, S_m \subset \mathcal{S}_n$ et $\dim(\mathcal{S}_n) := N_n \leq \sqrt{n/\log n}$.

Les hypothèses $(\mathcal{M}1)$–$(\mathcal{M}3)$ ne sont pas trop restrictives. Elles sont, en effet, vérifiées (cf. Barron et al. (1999)) pour les espaces F_{m_1} (et H_{m_2}) engendrés par : les bases trigonométriques, les bases de polynômes réguliers par morceaux, les bases d'ondelettes régulières, etc.

3 Résultats

3.1 Inégalité oracle

Soit une fonction h et un espace S, définissons

$$d(h,S) = \inf_{g\in S} \|h-g\| = \inf_{g\in S} \Big(\iint |h(x,y)-g(x,y)|^2 dxdy\Big)^{1/2}.$$

l'estimateur $\hat{\alpha}_{\hat{m}}$ satisfait l'inégalité oracle suivante.

Theorem 1 *Supposons que les hypothèses $(\mathcal{A}1)$ – $(\mathcal{A}5)$ et $(\mathcal{M}1)$ – $(\mathcal{M}3)$ sont vérifiées. Définissons la pénalité :*

$$\mathrm{pen}(m) := K_0(1+\|\alpha\|_{\infty,A})\frac{D_{m_1}D_{m_2}}{n}, \tag{6}$$

où K_0 est une constante numérique. Nous avons

$$\mathbb{E}(\|\alpha\mathbb{1}(A)-\hat{\alpha}_{\hat{m}}\|^2) \leq C \inf_{m\in\mathcal{M}_n}\{d^2(\alpha\mathbb{1}(A),S_m)+\mathrm{pen}(m)\} + \frac{C'}{n} \tag{7}$$

où $C = C(f_0, \|\alpha\mathbb{1}(A)\|^2)$ et C' sont des constantes dépendant de $\phi_1, \phi_2, \|\alpha\|_{\infty,A}, f_0$.

3.2 Upper bound for the rate

Nous pouvons obtenir avec le théorème précédent la vitesse de convergence de $\hat{\alpha}_{\hat{m}}$ sur des espaces de Besov anisotropes. Supposons que α (restreint à A) est dans l'espace de Besov anisotrope $B^{\boldsymbol{\beta}}_{2,\infty}(A)$ sur A de régularité $\boldsymbol{\beta} = (\beta_1, \beta_2)$ (cf. Triebel (2006)). Le résultat ci-dessous montre que $\hat{\alpha}_{\hat{m}}$ s'adapte à la régularité anisotrope inconnue de α.

Corollary 1 *Supposons que α restreint à A appartient à l'espace de Besov anisotrope $B^{\boldsymbol{\beta}}_{2,\infty}(A)$ sur A de régularité $\boldsymbol{\beta} = (\beta_1, \beta_2)$ tels que $\beta_1 > 1/2$ et $\beta_2 > 1/2$. Alors, sous les hypothèses du Théorème 1, nous obtenons*

$$\mathbb{E}\|\alpha - \hat{\alpha}_{\hat{m}}\|_A^2 = O(n^{-\frac{2\bar{\beta}}{2\bar{\beta}+2}}).$$

où $\bar{\beta}$ est la moyenne harmonique de β_1 et β_2 (i.e. $2/\bar{\beta} = 1/\beta_1 + 1/\beta_2$).

Nous montrons de plus que cette vitesse est optimale au sens minimax.

References

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. (1993) *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York.

Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113(3):301-413.

Beran, J. (1981) Nonparametric regression with randomly censored survival data. Technical report, Dept. Statist. Univ. California, Berkeley.

Dabrowska, D. M. (1987). Nonpametric regression with censored survival time data. *Scand. J. Statist.*, 14, 181-197.

Dabrowska, D. M. (1989). Uniform consistency of the kernel conditional Kaplan-Meier estimate. Ann. Statist. 17(3):1157-1167.

Grégoire G. 1993). Least squares cross-validation for counting processes intensities. *Scand. J. Statist.*, 20, 343-360.

Heuchenne, C. and Van Keilegom, I. (2007). Location estimation in nonparametric regression with censored data. J. Multiv. Anal., 98, 1558-1582.

Li, G. and Doss, H. (1995). An approach to nonparametric regression for life history data using local linear fitting. *Ann. Statist.*, 23(3):787-823.

McKeague, I. W., and Utikal, K. J. (1991). Inference for a nonlinear counting process regression model. *Ann. Statist.*, 18, 1172-1187.

Ramlau-Hausen, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.*, 11, 453-466.

Reynaud-Bouret, P. (2006). Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, 12(4), 633-661.

Stute, W. (1986). Conditional empirical processes. *Ann. Statist.*, 14(2), 638-647.

Triebel, H. (2006). *Theory of function spaces. III.* Monographs in Mathematics, 100. Birkhäuser Verlag, Basel, 2006.